

Written evidence submitted by Dr Huy Duong (IT Consultant at Private sector IT company)

COVID-19's effects 2020 A-level grades for Matthews Arnold School in Oxford

Summary

To assess how well Ofqual's standardisation will work for a real school, I have analysed 2017-2019 A-level data for Matthew Arnold School in Oxford (MAS). MAS is a non-selective state school with 1100-1200 students, so it is quite typical. The school was rated Good by Ofsted in 2018, 2013, 2008, so its standards are stable. It has been led by the same Head Teacher since 2006, who was awarded an MBE in 2019, so both the leadership and the running of the school are mature as well as stable. If the Exceptional Arrangements have any chance of providing accurate grades to students from non-selective state schools, it will be at those like MAS.

Data from Matthew Arnold School in Oxford (MAS) suggests that for A-levels the Exceptional Arrangements as published so far has virtually no chance of providing grades to the students in a way that satisfies the double criteria of being fair to the individuals and controlling grade inflation nationally. This problem affects every A-level subject at MAS, and is likely to affect most A-level subjects at hundreds of comparable schools across the country. The risk to the students is that fairness to the individuals might be sacrificed.

This analysis focuses on A and A* grades because between 2017 and 2019 MAS's A* rates vary by a factor of 3 and its A*/A rates vary by a factor of 2.

Year	Total A*s for all subjects	Total As For all subjects	Total entries	Percentage of entries achieving A*	Percentage of entries achieving A	Percentage of entries achieving A/A*
2019	19	65	303	6	21	28
2018	11	37	252	4	15	19
2017	27	51	201	13	15	39

Wide fluctuations in all A-level subjects

With the overall A* rates and A*/A rates fluctuating from year to year by factors of 3 and 2 respectively, it is not surprising that the fluctuation within each subject is even higher. The table below shows that this is true even in maths and the physical sciences, where MAS traditionally has strong performance and where the cohorts are relatively large for a school of this size.

Subject	Year	A*	A	Cohort size	Percentage achieving A*	Percentage achieving A	Percentage achieving A* over 3 years	Range of yearly A* percentage	Percentage achieving A over 3 years	Range of yearly A percentage
Biology	2019	1	9	34	3	26	4	0-6	25	13-42
	2018	2	4	32	6	13				
	2017	0	8	19	0	42				
Chemistry	2019	1	4	20	5	20	4	0-9	25	20-36
	2018	0	4	17	0	24				
	2017	1	4	11	9	36				
Maths	2019	7	8	37	19	22	17	12-19	27	22-32
	2018	3	8	25	12	32				
	2017	6	9	31	19	29				
Further Maths	2019	2	2	5	40	40	29	0-40	36	25-40
	2018	0	1	4	0	25				
	2017	2	2	5	40	40				
Physics	2019	3	6	19	16	32	18	10-33	26	20-32
	2018	1	2	10	10	20				
	2017	3	2	9	33	22				

For example,

- for Chemistry, the A* and A rates fluctuate from 0% to 9% and 20% to 36% respectively,
- for Physics, the A* and A rates fluctuate from 10% to 33% and 20% to 32% respectively,
- for Further Maths, which has smaller cohorts, the A* rates fluctuate from 0% to 40%.

These large fluctuations raise the question: how much will Ofqual allow the 2020 rates to differ from the 2017-2019 average ones?

Take Physics for example, where the three-year average A* rate is 18%. Would allowing 33% be too generous? Not really, because students and MAS achieved it in one out three years, and it cannot be reasonably asserted that this year's cohort cannot do as well or better. On the other hand, if 33% is allowed and the sample principle is applied nationally, that might well result in grade inflation. Will Ofqual sacrifice fairness to the individual students to serve the national goal of no grade inflation?

This question is even more acute for Further Maths. The three-year average A* rate is 29%, but in two of those three years 40% achieved A*s. Therefore there is a reasonable probability that this year's cohort can achieve A*s at 40% rate or even higher.

The social sciences face similar problems, e.g, with the A* rates for psychology fluctuating from 0% to 15%, the A rates for economics from 7% to 21% and the A rates for Economic fluctuating 0 to to 50%. None of these are niche subjects with a handful of students.

Subject	Year	A*	A	Cohort size	Percentage achieving A*	Percentage achieving A	Percentage achieving A* over 3 years	Range of yearly A* percentage	Percentage achieving A over 3 years	Range of yearly A percentage
Business Studies	2019	0	1	14	0	7	0	0-0	8	7-11
	2018	0	1	14	0	7				
	2017	0	1	9	0	11				
Economics	2019	0	4	19	0	21	2	0-7	13	7-21
	2018	0	1	14	0	7				
	2017	1	1	15	7	7				
History	2019	1	8	33	3	24	3	0-7	23	21-24
	2018	0	4	19	0	21				
	2017	1	3	14	7	21				
Psychology	2019	0	3	23	0	13	0	0-0	18	0-50
	2018	0	0	15	0	0				
	2017	0	6	12	0	50				
Sociology	2019	0	2	21	0	10	4	0-15	7	0-15
	2018	0	0	20	0	0				
	2017	2	2	13	15	15				

Take Psychology for example. While the difference in allowing an A rate of 7% or 15% seems to be only an 8% difference, that difference means doubling or halving the number of students achieving an A. If Ofqual allows an A rate of 7% it cannot be confident that it is not halving the rate that the 2020 cohort deserves. On the other hand, if it allows an A rate of 15% it cannot be confident that it is not doubling the rate that that cohort deserves either.

The arts and languages, traditionally with smaller cohorts, face even more intractable problems. For example, A* rates for French varied from 0% to 100%. Suppose two students are entered for French this year, and MAS submit that they both deserve A*. The historical data and the ranking do not give the exam board enough information to prescribe that one of both do not deserve A*.

Subject	Year	A*	A	Cohort size	Percentage achieving A*	Percentage achieving A	Percentage achieving A* over 3 years	Range of yearly A* percentage	Percentage achieving A over 3 years	Range of yearly A percentage
Art & Design Graphics	2019	0	0	6	0	0	0	0-0	0	0-0
	2018	N/A	N/A	N/A	N/A	N/A				
	2017	N/A	N/A	N/A	N/A	N/A				
Art and Design Photography	2019	0	0	6	0	0	5	0-33	0	0-0
	2018	0	0	11	0	0				
	2017	1	0	3	33	0				
Art and Design Textiles	2019	0	0	3	0	0	0	0-0	29	0-50
	2018	N/A	N/A	N/A	N/A	N/A				
	2017	0	2	4	0	50				
English Literature	2019	1	5	19	5	26	16	5-30	25	24-26
	2018	2	4	17	12	24				
	2017	6	5	20	30	25				
Film Studies	2019	0	0	6	0	0	9	0-50	9	0-25
	2018	0	1	12	0	8				
	2017	2	1	4	50	25				
Fine Art	2019	0	2	8	0	25	8	0-29	25	14-33
	2018	2	1	7	29	14				
	2017	0	3	9	0	33				
French	2019	1	0	1	100	0	20	0-100	40	0-67
	2018	0	2	3	0	67				
	2017	0	0	1	0	0				
Music	2019	1	0	3	33	0	27	20-33	9	0-20
	2018	1	1	5	20	20				
	2017	1	0	3	33	0				
Spanish	2019	1	4	6	17	67	17	0-50	50	0-67
	2018	0	2	4	0	50				
	2017	1	0	2	50	0				

Even in “Arts & Design Graphics” and “Arts & Design Textiles” where the A* rate is 0% without variation, that is still not enough for the exam board to look at that data and dictate that that this year the can’t be two talented art students in each subject who deserve A*.

This real dataset demonstrates that the Exceptional Arrangements as published so far cannot provide grades to MAS’s students in a way that satisfies the double criteria of being fair to the individuals and controlling grade inflation nationally. Given how typical MAS is, tens of thousands A-level students and teachers in hundreds of schools around the country are facing the same problems.

Statistical model has fundamental problems

A fundamental weakness with the principles behind the Exceptional Arrangements is that, at least for non-selective schools, the causal link between the school and the rates of A*s and As at A-level is relatively weak, and the dependence on the individual students is stronger. This can be seen by the wide variations not just from year to year, but also from grade to grade. Take Further Maths as an example:

Subject	Year	A*	A	B	C	D	E	U	Cohort size	Percentage achieving A*	Percentage achieving A
Further Maths	2019	2	2	0	0	0	1	0	5	40	40
	2018	0	1	1	2	0	0	0	4	0	25
	2017	2	2	0	0	1	0	0	5	40	40

If the causal link between the school and the grades achieved were stronger and the dependence on the individual students were weaker, the 2019 grades would be distributed more uniformly than 2 A*s, 2 As and 1 E; the same applies for the 2017 grades. Furthermore, the grades would be more similar from year to year. But neither is the case.

With weak links like that, a lot of historical data is needed to estimate MAS's true grade distribution in a subject. Sufficient data means not just large cohorts but a large number of students in each grade band and a reasonable number of years. As MAS's data shows, there simply isn't sufficient data: the number of students in each grade band is very small, especially for the grades near the top and bottom ends of the scale. Therefore it is not possible to estimate MAS's true grade distributions with any reasonable confidence and tolerance. No amount of refining details for any statistical model can get around that.

Even suppose hypothetically there were enough historical data to estimate MAS's true grade distribution in a subject well enough, a large 2020 cohort would still be needed in order to use that grade distribution with confidence. Think of tossing a coin with 50% probability of getting heads. With 200 tosses, it is likely that there will be close to 100 heads and 100 tails. However, it would be ill-advised and dictatorial for someone to assert that with 4 tosses you are only allowed to get 2 heads and 2 tails, and that if you get something different then you have probably cheated. No amount of refining details for any statistical model can get around that either.

Therefore, with the numbers A-levels at MAS and comparable schools, there isn't enough 2017-2019 data to make useful assertions about how many students are allowed to get each grade in a subject.

Can Ofqual deliver fairness and consistency?

Ofqual claims that its standardisation will bring consistency to grades. However, it is highly doubtful that it can do this for A-levels.

It is true that Ofqual and the exam boards have to perform standardisation every year. However, in previous years the standardisation was inter-centre: for each subject, an exam board would use nationwide data for the cohort being assessed to set the grade boundaries, and the same set of grade boundaries, i.e, the same standard would apply to every school in the country. This year the standardisation is primarily intra-centre: Ofqual and the exam boards will, in effect, set one set of grade boundaries per subject per school, using only that school's data, and a major component of that data, i.e, the 2017-2019 data, is not about the cohort being assessed. What the proposed standardisation does is to force the grade distribution for a subject at a school to be consistent with the grade distribution at that school in the past three years. However, due to the weak links between the grade distribution in the past three years and the current cohort's ability, there is very little to help making the 2020 grade boundaries consistent between different schools. So Ofqual's claim that its standardisation will bring consistency to grades is unlikely to be fulfilled for A-levels.

The proposed standardisation might be collectively fair, in the sense that the national grade distribution will be similar to previous years, which in turn is due to the fact that some students will get higher grades than they deserve, and some get lower. However, that doesn't mean that it is fair to the individuals. Furthermore, due to the high relative fluctuations in A* and A rates at non-selective schools, the standardisation process is more likely to award the talented students in these schools the wrong grades.

Appendix: 2017-2019 published A-level results from Matthew Arnold School in Oxford

	Year	A*	A	B	C	D	E	U	Cohort size	Percentage achieving A*	Percentage achieving A	Percentage achieving A* over 3 years	Range of yearly A* percentage	Percentage achieving A over 3 years	Range of yearly A percentage
Art & Design Graphics	2019	0	0	2	2	2	0	0	6	0	0	0	0-0	0	0-0
	2018	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A				
	2017	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A				
Art and Design Photography	2019	0	0	2	2	2	0	0	6	0	0	5	0-33	0	0-0
	2018	0	0	4	5	1	1	0	11	0	0				
	2017	1	0	2	0	0	0	0	3	33	0				
Art and Design Textiles	2019	0	0	0	2	1	0	0	3	0	0	0	0-0	29	0-50
	2018	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A				
	2017	0	2	1	1	0	0	0	4	0	50				
Biology	2019	1	9	7	5	6	5	1	34	3	26	4	0-6	25	13-42
	2018	2	4	2	10	9	5	0	32	6	13				
	2017	0	8	1	4	3	3	0	19	0	42				
Business Studies	2019	0	1	2	4	5	1	1	14	0	7	0	0-0	8	7-11
	2018	0	1	4	2	3	4	0	14	0	7				
	2017	0	1	3	3	1	0	1	9	0	11				
Chemistry	2019	1	4	6	3	3	1	2	20	5	20	4	0-9	25	20-36
	2018	0	4	3	4	5	0	1	17	0	24				
	2017	1	4	6	0	0	0	0	11	9	36				
Economics	2019	0	4	7	4	2	2	0	19	0	21	2	0-7	13	7-21
	2018	0	1	5	3	5	0	0	14	0	7				
	2017	1	1	8	4	1	0	0	15	7	7				
English Literature	2019	1	5	4	7	2	0	0	19	5	26	16	5-30	25	24-26

	2018	2	4	3	7	0	1	0	17	12	24				
	2017	6	5	3	4	2	0	0	20	30	25				
Film Studies	2019	0	0	2	3	1	0	0	6	0	0	9	0-50	9	0-25
	2018	0	1	4	3	3	1	0	12	0	8				
	2017	2	1	1	0	0	0	0	4	50	25				
Fine Art	2019	0	2	3	0	3	0	0	8	0	25	8	0-29	25	14-33
	2018	2	1	0	2	1	1	0	7	29	14				
	2017	0	3	5	1	0	0	0	9	0	33				
French	2019	1	0	0	0	0	0	0	1	100	0	20	0-100	40	0-67
	2018	0	2	0	0	1	0	0	3	0	67				
	2017	0	0	1	0	0	0	0	1	0	0				
Geography	2019	0	5	4	2	4	0	0	15	0	33	0	0-0	17	7-33
	2018	0	1	4	6	2	1	0	14	0	7				
	2017	0	1	5	7	0	0	0	13	0	8				
History	2019	1	8	14	7	3	0	0	33	3	24	3	0-7	23	21-24
	2018	0	4	7	6	2	0	0	19	0	21				
	2017	1	3	7	3	0	0	0	14	7	21				
Maths	2019	7	8	7	5	4	3	3	37	19	22	17	12-19	27	22-32
	2018	3	8	8	2	4	0	0	25	12	32				
	2017	6	9	9	2	2	2	1	31	19	29				
Further Maths	2019	2	2	0	0	0	1	0	5	40	40	29	0-40	36	25-40
	2018	0	1	1	2	0	0	0	4	0	25				
	2017	2	2	0	0	1	0	0	5	40	40				
Music	2019	1	0	0	1	1	0	0	3	33	0	27	20-33	9	0-20
	2018	1	1	2	1	0	0	0	5	20	20				
	2017	1	0	0	2	0	0	0	3	33	0				

Physics	2019	3	6	6	3	0	1	0	19	16	32	18	10-33	26	20-32
	2018	1	2	5	1	1	0	0	10	10	20				
	2017	3	2	2	2	0	0	0	9	33	22				
Psychology	2019	0	3	6	7	3	4	0	23	0	13	0	0-0	18	0-50
	2018	0	0	2	5	7	1	0	15	0	0				
	2017	0	6	4	2	0	0	0	12	0	50				
Sociology	2019	0	2	6	8	5	0	0	21	0	10	4	0-15	7	0-15
	2018	0	0	1	9	5	4	1	20	0	0				
	2017	2	2	6	3	0	0	0	13	15	15				
Spanish	2019	1	4	0	1	0	0	0	6	17	67	17	0-50	50	0-67
	2018	0	2	1	0	1	0	0	4	0	50				
	2017	1	0	1	0	0	0	0	2	50	0				
Sport/PE Studies	2019	0	2	1	1	0	1	0	5	0	40	0	0-0	17	0-40
	2018	0	0	0	4	4	1	0	9	0	0				
	2017	0	1	0	0	2	1	0	4	0	25				

July 2020